# A comparison of the cognitive difficulties posed by SPARQL query constructs

# A comparison of the cognitive difficulties posed by SPARQL query constructs

Paul Warren[1[0000-0002-4209-1436]] and Paul Mulholland[1[0000-0001-6598-0757]]
[1] Knowledge Media Institute, The Open University, Milton Keynes, MK7 6AA, U.K.
{paul.warren, paul.mulholland}@open.ac.uk

**Abstract.** This study investigated difficulties in the comprehension of SPARQL. In particular, it compared the declarative and navigational styles present in the language, and various operators used in SPARQL property paths. The study involved participants selecting possible answers given a SPARQL query and knowledgebase. In general, no significant differences were found in terms of the response time and accuracy with which participants could answer questions expressed in either a declarative or navigational form. However, UNION did take significantly longer to comprehend than both braces and vertical line in property paths; with braces being faster than vertical line. Inversion and negated property paths both proved difficult, with their combination being very difficult indeed. Questions involving MINUS were answered more accurately than those involving negation in property paths, in particular where predicates were inverted. Both involve negation, but the semantics are different. With the MINUS questions, negation and inversion can be considered separately; with property paths, negation and inversion need to be considered together. Participants generally expressed a preference for data represented graphically, and this preference was significantly correlated with accuracy of comprehension. Implications for the design and use of query languages are discussed.

**Keywords:** SPARQL, user experience, participant study.

## 1 Introduction

The original specification of the SPARQL query language, SPARQL1.0 [1], employed a declarative syntax style, heavily influenced by SQL. Subsequently, SPARQL1.1 [2] introduced a number of new features, including a navigational syntax using *property paths*. This syntax was based on regular expressions and enabled the more compact expression of certain queries, besides the ability to define chains of unbounded length. The goal of the study reported here was to compare the ease of comprehension of the declarative and navigational styles, and to investigate the difficulties which people have with some of the property path features. The motivation for the work was to advise on the writing of easily intelligible queries; and to make recommendations for the future development of SPARQL and similar languages. The knowledgebases used in the study were expressed textually and graphically, and this also enabled a comparison of participants' reaction to the two formats. We used comprehension tasks because comprehension is fundamental to creating and sharing queries, and to interpreting the results of queries. A study such as this could usefully be complemented by a study involving query creation.

Section 2 reviews related work. Section 3 lists those features of the language which were used in the study, and describes the study's specific objectives. Section 4 describes how the study was organized. Sections 5 to 8 then describe each of the four study sections and present their results. Section 9 reports on what influence the prior knowledge of the participants had on their responses. Section 10 discusses the participant's usage of the textual and graphical forms of the knowledgebases. Finally, Section 11 summarizes the main findings and makes some recommendations.

## 2 Related work

A number of researchers have analysed query logs from RDF data sources. Gallego et al. [3], and Rietveld and Hoekstra[4] looked at the frequency of use of various SPARQL features. Of relevance to this study, they found that UNION was among the more frequently used features. More recently, Bielefeldt et al [5] have found appreciable usage of property path expressions. Bonifati et al. [6] looked at the relative usage of property path features. They found that negated property sets (!), disjunction (|), zero or more (*) and concatenation (/) were relatively frequently used. Complementing these studies, Warren and Mulholland [7] have surveyed the usage of SPARQL1.1 features. They report that 71% of their respondents used property paths. Similarly to Bonifati et al. [6], Warren and Mulholland [7] found that /, * and | were relatively frequently used operators. They also found that one or more (+) was relatively frequently used, and that ^ and ? were also used to a certain extent. However, ! was little used. By contrast, there has been little work reported on the user experience of query languages. There were a number of studies in the early days of database query languages, e.g. see Reisner [8]. More recently, there have been some studies of the usability of certain semantic web languages, e.g. Sarker et al. [9] have investigated rule-based OWL modelling and Warren et al. [10] have investigated Manchester OWL Syntax. However, to the authors' knowledge, there have been no studies investigating the usability of semantic web query languages.

## 3 SPARQL – declarative and navigational

The study made use of the following declarative features of the language: join, represented by a dot; UNION; and MINUS, i.e. set difference[1]. The property path features used were: concatenation (/); disjunction (|); inverse (^); negated property sets (!); and one or more occurrences of an element (+). We also used the braces notation, where, *{m,n}* after a path element implies that the element occurs at least *m*, and no more than *n* times. In fact, the braces notation was not included in the final W3C recommendation for SPARQL1.1. However, this notation was present in a working draft for SPARQL1.1 property paths [11], and is implemented in the Apache Jena Fuseki

[1] Although part of the language's declarative style, MINUS was introduced in SPARQL1.1.

SPARQL server[2]. Moreover, the braces notation has been suggested for introduction in the next SPARQL standard[3]. Additionally, the SELECT and WHERE keywords were used. The use of these features is illustrated in Sections 5 to 8. The specific objectives of the study were to:

- compare the original declarative syntax style used in SPARQL1.0 with the navigational style introduced in SPARQL1.1 (see Section 5);
- compare the use of braces, vertical line and plus in property paths; and compare these property path constructs with the use of UNION (see Section 6);
- investigate the understanding of inversion and negation in property paths (see Sections 7 and 8).

Considering the last of these points, the study also considered the use of MINUS. This is another way of introducing negation into queries, albeit with a different semantics to that of negation in property paths. As described in Section 7, the study was able to compare how people reasoned about negation in the two cases.

## 4 Organization of the study

The study was conducted on an individual basis, on the experimenter's laptop. The MediaLab application[4] was used to collect responses and record response times. There were 20 questions, divided over four sections. Each question displayed a small knowledgebase, shown on the left of the screen as a set of triples, and on the right diagrammatically. For each section, all the questions used the same knowledgebase, displayed in the same way. The screen also displayed a SPARQL query. This was in a simplified version of the language, in particular without any reference to namespaces. Finally, there were four possible solutions to the query. Participants were required to tick which of the four solutions were valid. It was made clear that the number of valid solutions could range between zero and four inclusive. Participants could then click on *Continue* at the bottom right to move on to the next question. MediaLab recorded the response or lack of response to each solution, and the time for the question overall. Figure 1 shows a sample screen, in this case for one of the questions in Section 5. For all screenshots see: https://doi.org/10.21954/ou.rd.11931645.v1 .

---

[2] https://jena.apache.org/documentation/fuseki2/

[3] See https://github.com/w3c/sparql-12/issues/101. The likely reason for braces not being included in SPARQL1.1 property paths is the difficulty in deciding whether to opt for counting (bag) or non-counting (set) semantics. The former was the default in the original SPARQL standard. However, after the discovery of possible performance issues (see [12]), non-counting semantics were introduced in SPARQL1.1 specifically for property paths of unlimited length, i.e. using star (*) or plus (+); while leaving counting semantics as the default for all other SPARQL constructs.

[4] Provided by Empirisoft: http://www.empirisoft.com

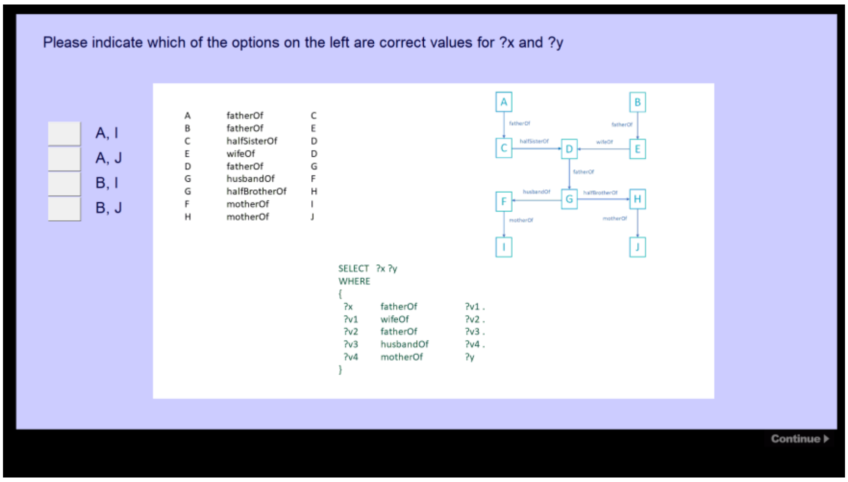



**Fig. 1.** Example question screen. This screen is for one of the questions discussed in Section 5.

Before the study the participants were presented with a six-page handout which described all the SPARQL features used in the study. Participants were asked to read the handout before beginning the study and encouraged to refer to the handout whenever necessary when completing the study. At the beginning of the study there were two screens providing more information about the study, and then a number of screens asking the participants for information on their knowledge of SPARQL, SQL, or any other query language, see Section 9. There was then a practice question, designed to introduce the participants to the format of the study; data from this question was not used in the analysis. For this question, and for this question only, the solution was subsequently presented to the participants. Participants then worked through the four sections. The order of presentation of the sections, and of the questions within the sections, were randomized. Randomization of the order of the sections mitigated the chance that performance might vary between the sections, e.g. the first because of unfamiliarity with the format of the question and the last because of fatigue. Randomization of the order of the questions similarly mitigated these effects, and also any learning effects between questions.

Participants were recruited from the authors' own institute and from a variety of other research and industrial environments. They were either computer scientists, with or without a knowledge of SPARQL, or else workers in other disciplines who made use of SPARQL. After a pilot with one participant, the study involved 19 participants, of whom 6 were female. The study was a within-participants study, so that between participant variability would equally affect all conditions. The research was approved by the Open University Human Research Ethics Committee (HREC/3175) and all participants signed a consent form prior to taking part. The study took place during March, April and May 2019.

The analysis was based on the accuracy and the response time. In each section below, accuracy is shown as the percentage of correct responses for each putative query solution. Comparisons of accuracy used logistic analysis of deviance, i.e. assumed a binomial distribution of correct / not correct responses to each putative query solution. Response time data were collected per question. Analysis of the response time data indicated that they were positively skewed and hence did not follow a normal distribution. For this reason, non-parametric tests have been used in analyzing the response time data. Because this was a within-participants study, where appropriate these non-parametric tests are paired tests. All statistical analysis used the R statistical package [13] . Throughout, $p < 0.05$ was taken as indicating statistical significance. Where pairwise analysis was undertaken, corrections were made for multiple testing.

## 5 Declarative versus property path, forward versus inverse

**Questions** The four questions in this section were concerned with comparing participant performance on the declarative and property path syntactic styles, and on forward and inverse predicates. Figure 1 shows the question in the declarative style and using forward predicates. There is only one valid solution: (B, I). There was an analogous question, with the same solution, using a property path in the WHERE clause:

- {?x fatherOf / wifeOf / fatherOf / husbandOf / motherOf ?y}

The other two questions were in the declarative and property path styles, using inverse predicates, with the following WHERE clauses:

- {?x ^motherOf ?v1 . ?v1 ^halfBrotherOf ?v2 . ?v2 ^fatherOf ?v3 . ?v3 ^halfSisterOf ?v4 . ?v4 ^fatherOf ?y}
- {?x ^motherOf / ^halfBrotherOf / ^fatherOf / ^halfSisterOf / ^fatherOf ?y

The proposed solutions for these two inverse predicate questions were (from top to bottom): (I, A), (I, B), (J, A), (J, B); (J, A) is the correct solution.

The four questions were designed so that, considering the diagram in Figure 1, the correct solution for the two questions with forward predicates required a traversal from top right to bottom left; whereas for the other two questions with inverse predicates, a traversal from bottom right to top left was required. Thus, each of the four queries made similar traversals of the knowledgebase, to enable a meaningful comparison.

**Results** Table 1 shows the percentage of correct responses, for each of the proposed solutions and overall for each question, besides the mean and standard deviation times for each question. In the table, and in subsequent similar tables, valid solutions are identified by showing their percentage of correct responses underlined and in bold. A two-factor analysis of deviance indicated a significant difference in accuracy between forward and inverse predicates ($p = 0.012$), but no significant difference between the declarative and property path styles ($p = 0.406$) and no interaction effect ($p = 0.947$). Paired Wilcoxon tests[5] showed a significant difference in response time between the forward and inverse predicates ($p = 0.0003$) but not between the two

[5] The Wilcoxon test is a non-parametric test used in a within-participants study to compare two conditions. It can be considered as a non-parametric analogue of a paired t-test.

syntactic styles (p = 0.405). When the questions in the two styles were analyzed separately there was a significant difference in response time between the forward and inverse predicates for both the declarative questions (p = 0.023) and the property path questions (p = 0.005).

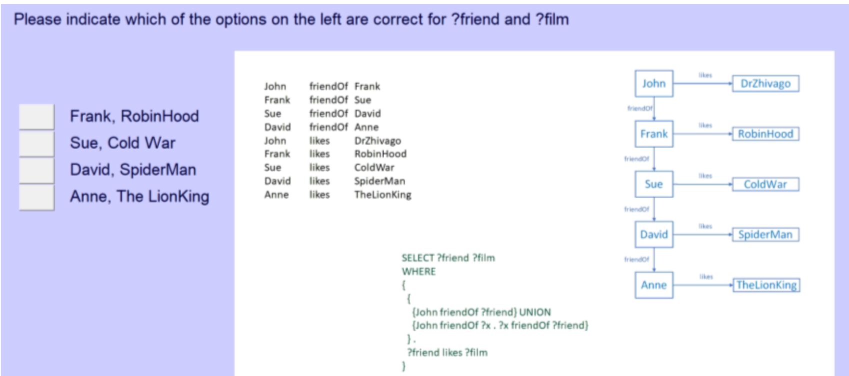
**Table 1.** Data for ‘declarative versus property path, forward versus inverse’ questions

| **predicate direction** | **syntax** | **Percentage correct** | | | | | **mean time (secs)** | **s.d. (secs)** |
|---|---|---|---|---|---|---|---|---|
| | | (A, I) | (A, J) | (B, I) | (B, J) | overall | | |
| forward | declarative | 100% | 94.7% | **94.7%** | 94.7% | 96.1% | 75.7[6] | 85.2 |
| | property path | 94.7% | 94.7% | **94.7%** | 94.7% | 94.7% | 48.1 | 29.2 |
| inverse | | (I, A) | (I, B) | (J, A) | (J, B) | overall | | |
| | declarative | 100% | 89.5% | **84.2%** | 84.2% | 89.5% | 108.0 | 88.6 |
| | property path | 84.2% | 84.2% | **73.7%** | 100% | 85.5% | 114.3 | 103.3 |

**Discussion** Participants answered the questions with inverse predicates less accurately and they took longer to do so. Inversion can be seen as cognitively analogous to negation, which has been extensively studied, e.g. [14], [15]. They both require the construction of an initial mental model, which must then be inverted or negated, a process which both takes time and increases the probability of error.

## 6 Disjunction

**Questions** This section of the study was concerned with comparing four ways of achieving disjunction in a query: using the UNION keyword in the declarative style; or using braces, vertical line or plus in the property path style. Figure 2 displays a portion of the screenshot for a question using the UNION keyword, showing the knowledgebase used for each question in the section.




6 This time is increased by an outlier of 378 secs. When this is removed, the time is 58.9 secs.

**Fig. 2.** Part of a question screen from the disjunction section.

There were also two analogous questions, using the vertical line and braces notations:

- {John friendOf | (friendOf / friendOf) ?friend . ?friend likes ?film}
- {John friendOf{1, 2} ?friend . ?friend likes ?film}

All three questions had the same proposed solutions, as shown in Figure 2. Of these, (Frank, RobinHood) and (Sue, ColdWar) were valid. There were another three analogous questions using UNION, vertical line and braces, extending the 'reach' of the friendOf chain to four steps:

- {{{John friendOf ?friend} UNION
  {John friendOf ?x . ?x friendOf ?friend} UNION
  {John friendOf ?x . ?x friendOf ?y . ?y friendOf ?friend} UNION
  {John friendOf ?x . ?x friendOf ?y . ?y friendOf ?z . ?z friendOf ?friend}} .
  ?friend likes ?film}
- {John friendOf | (friendOf / friendOf) | (friendOf / friendOf / friendOf)
  | (friendOf / friendOf / friendOf / friendOf) ?friend . ?friend likes ?film}
- {John friendOf{1, 4} ?friend . ?friend likes ?film}

The same proposed solutions were used as in Figure 2, and this time all were valid. Thus, there were six questions comparing UNION, vertical line and braces at what might be considered two levels of complexity, i.e. {1, 2} and {1, 4}. Thus, for each level of complexity, the three questions had the same solutions.

Finally, there was a seventh question, employing plus:

- {John friendOf+ ?friend . ?friend likes ?film}

For this question the topmost proposed solution, i.e. (Frank, RobinHood) was replaced with (John, DrZhivago); this was to test understanding of the plus operator. Thus, this topmost solution was not valid, whilst the remaining three were valid.

**Results** Table 2 shows the data for each of the seven questions, with column headings identifying the proposed solutions by the value bound to *?friend*. Consider, first, the six questions excluding plus. A two-way analysis of deviance revealed that the accuracy of response for the three questions with reach 2 was significantly better than that for the three questions with reach 4 ($p = 0.009$), whilst there was no significant difference between the three operators ($p = 0.986$) and no interaction effect ($p = 0.297$). Turning to the response time data, a Wilcoxon test indicated no significant difference between the questions with reach 2 and reach 4 ($p = 0.769$). A Friedman test did indicate a significant difference between the three operators ($p = 0.0001$). In fact, a pairwise Wilcoxon test indicated a significant difference for each comparison (brace:union, $p = 0.0001$; brace:vertical, $p = 0.021$; vertical:union, $p = 0.030$). Friedman[7] tests also revealed a significant difference in response time between the operators at both levels of complexity (reach 2: p = 0.016; reach 4: p = 0.004).

**Table 2.** Data for the disjunctive questions.
N.B. * indicates not significantly greater than chance (one-sided test)

| reach | operator | Percentage correct | | | | | mean time (secs) | s.d. (secs) |
|---|---|---|---|---|---|---|---|---|
| | | Frank | Sue | David | Anne | overall | | |
| 2 | UNION | **84.2%** | **89.5%** | 89.5% | 94.7% | 89.5% | 95.3 | 86.3 |
| | vert. line | **78.9%** | **68.4%*** | 84.2% | 100% | 82.9% | 76.2 | 70.7 |
| | braces | **78.9%** | **73.7%** | 84.2% | 100% | 84.2% | 53.0 | 68.1 |
| 4 | UNION | **68.4%*** | **63.2%*** | **63.2%*** | **94.7%** | 72.4% | 82.0 | 52.5 |
| | vert. line | **84.2%** | **73.7%** | **73.7%** | **84.2%** | 78.9% | 66.7 | 51.3 |
| | braces | **78.9%** | **68.4%*** | **68.4%*** | **89.5%** | 76.3% | 40.1 | 30.5 |
| | | John | Sue | David | Anne | | | |
| ∞ | plus | 94.7% | **68.4%** | **68.4%** | **73.7%** | 76.3% | 73.8 | 39.2 |

A comparison was also made between the three questions with reach 4 and the question employing plus. For these questions, the analysis of accuracy excluded the topmost solution (i.e. leftmost in Table 2), which was different for the question with plus, i.e. it was based on the three solutions common to all four questions, which were all valid. On this basis, an analysis of variance indicated no difference in accuracy of response between the questions (p = 0.851). For the response time data, a Friedman test indicated a significant difference between the four operators (p = 0.006).

**Discussion** The analysis of the six questions with UNION, vertical line and braces, indicated that the queries with longer reach were less accurately answered. On the other hand, the difference in reach made no significant difference to the response times. Conversely, the choice of operator made no significant difference to accuracy but did make a significant difference to response times, with the braces operator being significantly faster than the other operators. When the plus operator was included in the analysis, there was again a significant difference in response time but not in accuracy between the operators. The speed of interpretation of the braces operator may be due to the clarity of expression it permits, avoiding the combinatorial explosion which occurs with UNION and vertical line. The plus operator permits the same clarity of expression, but unlike the braces notation, its meaning is not explicit. The longer time for the plus operator, compared with the braces, may also be due to the difference in the sets of solutions.

---

[7] The Wilcoxon test is a non-parametric test used in a within participants study to compare more than two conditions. It can be regarded as a non-parametric analogue of a repeated measures ANOVA.

## 7 MINUS and negated property sets

**Questions** This section contained six questions employing two forms of negation introduced into SPARQL1.1: MINUS and negated property sets. In each case there were three questions: with forward predicates, inverse predicates and a disjunction of forward and inverse predicates. The questions were designed to examine participants' reasoning with negation, with negation when combined with an inverse predicate; and also to compare participants' treatment of negation in negated property sets and in constructs with MINUS. Figure 3 displays the screenshot for the MINUS question with forward predicate. The other two MINUS questions have WHERE clauses:

- {?x ^teacherOf ?y} MINUS {?x ^parentOf ?y}
- {?x teacherOf | ^teacherOf ?y} MINUS {?x parentOf | ^parentOf ?y}

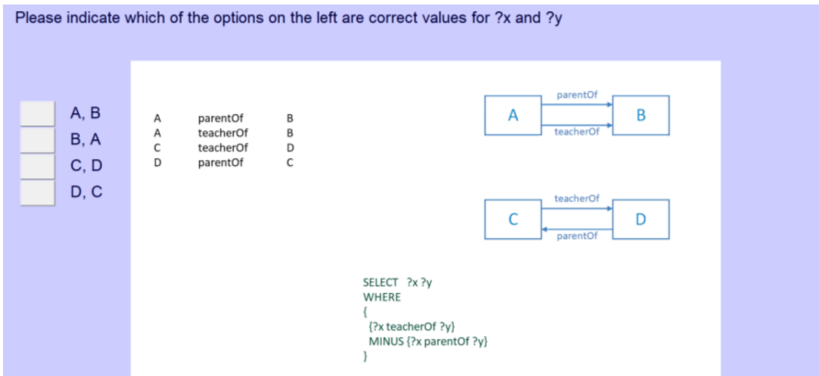



**Fig. 3.** Part of a question screen from the MINUS and negated property set section.

The three questions with negated property sets have WHERE clauses:

- {?x !parentOf ?y}
- {?x !^parentOf ?y}
- {?x !(parentOf | ^parentOf )?y}

The first of these three is satisfied by using any predicate from *?x* to *?y*, other than *parentOf*. The second is satisfied by using any inverse predicate from *?x* to *?y*, other than the inverse of *parentOf*, i.e. it is equivalent to *?y !parentOf ?x*. The third is satisfied by using any forward or inverse predicate from *?x* to *?y* other than *parentOf* and its inverse, i.e. it is equivalent to *{?x !parentOf ?y} UNION {?y !parentOf ?x}*.

All six questions use the same knowledgebase and the same set of proposed solutions. It is important, however, to note that the semantics of MINUS and negated property sets are different. This can be seen in Table 3, which shows the data for this section and indicates the valid solutions, by showing the percentage of correct responses for these solutions underlined and in bold. For the question with MINUS and a disjunction of forward and inverse predicates, there are no valid solutions; whilst for the corresponding question with a negated property set, all the solutions are valid.

**Table 3.** Data for MINUS and negated property set questions
N.B. * not significantly greater than chance; † significantly less than chance (one-sided tests)

| form of negation | predicate direction | Percentage correct | | | | | mean time (secs) | s.d. (secs) |
|---|---|---|---|---|---|---|---|---|
| | | A, B | B, A | C, D | D, C | overall | | |
| MINUS | forward | 94.7% | 94.7% | **78.9%** | 84.2% | 88.2% | 50.4 | 30.2 |
| | inverse | 84.2% | 78.9% | 73.7% | **52.6%*** | 72.4% | 93.2 | 52.4 |
| | disjunction | 84.2% | 73.7% | 73.7% | 84.2% | 78.9% | 146.1 | 133.4 |
| negated property sets | forward | **57.9%*** | 68.4%* | **89.5%** | 89.5% | 76.3% | 54.7 | 43.9 |
| | inverse | 31.6% | **31.6%** | 31.6% | **47.4%** | 35.5%† | 82.7 | 51.3 |
| | disjunction | **47.4%** | **36.8%** | **52.6%** | **26.3%†** | 40.8% | 86.8 | 83.8 |

**Results** A two-way analysis of deviance indicated a significant difference in accuracy between the MINUS and negated property path questions ($p = 10^{-11}$) and between the three uses of predicates ($p = 3 \times 10^{-8}$), with no interaction effect ($p = 0.311$). When the MINUS questions are considered separately, a one-way analysis of deviance indicated a significant difference in accuracy between the three uses of predicates ($p = 0.046$). A subsequent Tukey HSD analysis indicated a significant difference between the forward and inverse usages ($p = 0.045$), but not between forward and disjunction ($p = 0.283$) and inverse and disjunction ($p = 0.612$). A similar analysis for the negated property set questions again indicated a significant difference between the predicate usages ($p = 2 \times 10^{-7}$). A Tukey HSD also indicated a significant difference between forward and inverse usages ($p = 10^{-4}$) and in this case between forward and disjunction ($p = 10^{-4}$), but not between inverse and disjunction ($p = 0.782$). The analysis indicates that the operator usage does have a significant effect on accuracy, and that effect is more extreme for the negated property set questions. It is particularly noteworthy that for the negated property set question employing only an inverse predicate, all the proposed solutions were responded to less accurately than chance; and for the negated property set question employing a disjunction of forward and inverse predicates, three of the four solutions were responded to less accurately than chance.

Turning to the response time, a Wilcoxon test showed no significant difference between the MINUS and the negated property set questions ($p = 0.075$), but a Friedman test showed a significant effect of predicate usage ($p = 0.006$). However, this latter effect appears to originate from the MINUS questions. Considering the MINUS questions separately, there was a significant effect of predicate usage on response time ($p = 0.006$). Considering the property path questions separately, there was no significant effect of predicate usage ($p = 0.331$). For the MINUS questions, pairwise Wilcoxon tests indicated a significant difference between forward and inverse ($p = 0.005$), and between forward and disjunction ($p = 0.0002$), but not between inverse and disjunction ($p = 0.087$).

**Discussion** The MINUS questions were answered significantly more accurately than the negated property set questions, but there was no significant difference in response times between the two sets of questions. The predicate usage had an effect on accuracy for both sets of questions, but on response time only for the MINUS questions. We can consider how the participants may be formulating answers to these questions. For the MINUS questions, participants are required to compute two sets and then find the set difference. For both sets they are required to think in terms of forward predicates, inverse predicates, and both forward and inverse predicates. Thinking in terms of inverse predicates is less accurate and slower than for forward predicates, because of the need to perform the inversion operations. Note that the mean time for the disjunction of forward and inverse predicates is close to the sum of the mean times in the other two cases. In part this effect is a chance effect occurring in aggregate. However, at the participant level, the response time for the question employing both forms of the predicate is generally relatively close to the sum of the times for the other two questions. This suggests that participants formed each of the two sets by considering the forward and inverse predicates separately.

For the MINUS questions, forming the set difference comes after the creation of the two sets, i.e. it is not required to consider negation and inverse at the same time. For the negated property set questions, the question using an inverse predicate and the question using the disjunction of forward and inverse predicates, require that negation and inverse be considered at the same time. This is likely to be the reason why, when using negated property sets, the inverse and disjunction questions were answered significantly less accurately than the forward question. For the negated property set questions, one might expect the disjunction of forward and inverse predicates to be answered less accurately than the question with solely an inverse predicate, since the former requires the manipulation of two mental models. However, for the question with a negated property set and the disjunction of predicate usages, all the solutions were valid. Two of the solutions, (A, B) and (C, D), required usage of the forward predicate, and these were the two where participants performed best; although not as well as for the question with forward predicate alone. The other two solutions required use of the inverse predicates, and here participants were less accurate. As a result, for the property path questions, the accuracy of the disjunctive question was between the other two.

## 8 Negated property sets and braces

**Questions** This section further examined the difficulties of negated property sets, in a situation where the braces notation was used. The questions were designed to examine participants' reasoning with negated property sets and inverse predicates in a more complex use than that for the questions in Section 7; in particular where a chain of predicates had to be considered. Figure 4 shows one of the questions, in this case using a forward predicate.

There were two other questions, using an inverse predicate and a disjunction of forward and inverse predicates:

- {?x (!^parentOf){1, 3} ?y}
- {?x (!(parentOf | ^parentOf)){1, 3} ?y}

All three questions used the same knowledgebase and the same proposed solutions.

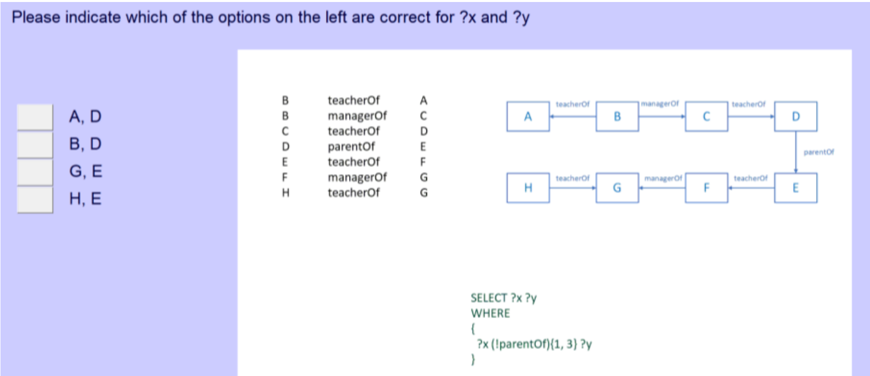



**Fig. 4.** Part of a question screen from the negated property set and braces section.

**Results** Table 4 shows the data for this section. A one-factor analysis of deviance indicated no significant difference in accuracy between the three questions (p = 0.293). A two-factor analysis, with both question and solution as factors indicated no significant difference for question (p = 0.292) or for solution (p = 0.851), but did indicate a significant interaction effect (p = 0.006). This is consistent with the data in Table 4, where it can be seen that the between-question variation is much greater for some solutions than it is overall. Turning to the response time data, a Friedman test indicates that there is no significant difference between the response times for the three cases (p = 0.368). The large mean time for the inverse predicate question is largely due to three response times of over 300 seconds each. The presence of these outliers is suggested by the large standard deviation. When they are removed, the mean time reduces to 90.4 seconds.

**Table 4.** Data for negated property set and braces questions
N.B. * not significantly greater than chance; † significantly less than chance (one-sided tests)

| **predicate direction** | **Percentage correct** | | | | | **mean time (secs)** | **s.d. (secs)** |
|---|---|---|---|---|---|---|---|
| | A, D | B, D | G, E | H, E | overall | | |
| forward | 63.2%* | **84.2%** | 57.9%* | 57.9%* | 65.8% | 113.5 | 78.1 |
| inverse | 68.4%* | 26.3%† | **57.9%*** | 63.2% | 53.9%* | 144.8 | 144.6 |
| disjunction | **52.6%*** | **78.9%** | **73.7%** | **47.4%** | 63.2% | 96.9 | 75.7 |
| all questions | 61.4%* | 63.2%* | 63.2%* | 56.1%* | 61.0% | | |

**Discussion** In the previous section, considering the negated property set questions, the valid solutions for the disjunctive query were the union of the valid solutions for the other two queries. In this section, however, the valid solutions for the disjunctive query include two solutions, (A, D) and (H, E), which are not valid solutions for either of the other two questions. This arises because, unlike the other two solutions and the solutions for the disjunctive question in the previous section, these two solutions make use of a combination of forward and inverse predicates. For these solutions, participants needed to consider both forward and inverse predicates. This may explain why accuracy for these two solutions was less than for the other two. It is also noticeable that, for the inverse question, accuracy for the solution (B, D) was significantly less than chance. This is the solution which is valid for the forward predicate, so it seems likely that some participants were not inverting the predicate, and simply treating the question as they would the forward predicate question.

## 9 Effect of prior participant knowledge

At the beginning of the study, participants were asked to rate their knowledge of SPARQL, SQL, and any other query language on a four-category scale. Table 5 shows the distribution of responses for each of the three questions.

**Table 5.** Expertise in query languages (percentage participants)

| N = 19 | **no knowledge at all** | **a little knowledge** | **some knowledge** | **expert knowledge** |
|---|---|---|---|---|
| **SPARQL** | 36.8% | 36.8% | 21.1% | 5.3% |
| **SQL** | 21.1% | 15.8% | 42.1% | 21.1% |
| **other query lang.** | 73.7% | 5.3% | 15.8% | 5.3% |

**Table 6.** Accuracy and mean response time per SPARQL expertise category

| | **no knowledge at all** | **a little knowledge** | **some knowledge** | **expert knowledge** |
|---|---|---|---|---|
| %age correct | 75.4% | 69.6% | 80.6% | 88.8% |
| mean time; s.d. (secs) | 104.1; 37.8 | 78.2; 32.3 | 75.5; 16.3 | 39.6; NA |

**Table 7.** Accuracy and mean response time per SQL expertise category

| | **no knowledge at all** | **a little knowledge** | **some knowledge** | **expert knowledge** |
|---|---|---|---|---|
| %age correct | 74.4% | 66.3% | 74.4% | 83.8% |
| mean time; s.d. (secs) | 96.5; 41.8 | 54.3; 11.0 | 87.9; 30.7 | 91.3; 42.0 |

Tables 6 and 7 show the percentage of correct responses to solutions, over all the 80 proposed solutions, and the mean response time per question, for each of the categories of expertise in SPARQL and SQL. A one-sided Spearman's rank test indicated that accuracy did not significantly correlate with prior knowledge of SPARQL (rho = 0.12, p = 0.306)[8], or with knowledge of SQL (rho = 0.25, p = 0.148). However, the mean response time did significantly correlate with knowledge of SPARQL (rho = -0.44, p = 0.031), but not with knowledge of SQL (rho = 0.09, p = 0.358). Finally, it was thought that performance might depend on the overall knowledge of query lan-

8 Spearman's rank correlation is a non-parametric measure of the correlation between the ranks of two variables. In this and subsequent Spearman's rank tests, the exact p-value could not be computed because of ties.

guages, represented by highest level of expertise for each participant over the three questions. However, there was no significant correlation with accuracy (rho = 0.35, p = 0.071) or with response time (rho = -0.02, p = 0.462). In summary, the only significant effect of prior knowledge is that knowledge of SPARQL reduced response time, possibly because participants familiar with SPARQL spent less time referring to the handout.

## 10 Textual and graphical representations

At the end of the study, participants were asked to describe their usage of the textual and graphical representations, according to the five categories shown in Table 8. These categories are arranged on an ordinal scale, going from an entirely textual approach at the top of the table, to an entirely graphical approach at the bottom. The table shows the percentage of participants in each of the categories, the percentage of correct responses to the proposed solutions, and the mean response time per question.

**Table 8.** Usage of textual and graphical information

| | **%age of respondents (N = 19)** | **%age correct responses** | **mean response time (secs)** | **s.d. (secs)** |
|---|---|---|---|---|
| I used only the textual information. | 5.3% | 61.3% | 97.3 | NA |
| I used mostly the textual information, but also made some use of the graphical information. | 5.3% | 67.5% | 113.8 | NA |
| I used the textual and graphical information about equally. | 10.5% | 69.4% | 74.4 | 49.3 |
| I used mostly the graphical information, but also made some use of the textual information. | 57.9% | 72.0% | 87.0 | 40.9 |
| I used only the graphical information. | 21.1% | 91.6% | 75.1 | 7.9 |

A two-sided Spearman's rank test showed that preference for the graphical representation correlated significantly with accuracy of response (rho = 0.50, p = 0.029). Thus, not only did the majority of participants prefer the graphical representation, but this preference correlated with increased accuracy. There was, however, no significant correlation with response time (rho = -0.11, p = 0.651). The questions in Section 5 permit a comparison of how this effect differs between the two styles. For the navigational style there was a significant correlation between preference for graphics and accuracy (rho = 0.56, p = 0.013); this was not the case for the declarative style (rho = 0.34, p = 0.161). For neither of the two styles was there a significant correlation with time (navigational: rho = -0.28, p = 0.241; declarative: rho = -0.38, p = 0.108).

## 11 Summary and recommendations

Earlier in the paper we set out three goals. Firstly, we wanted to investigate whether there was any difference in the comprehension of questions in declarative and navigational form. Section 5 indicated that in general there was no significant difference in the styles. However, Section 6 does demonstrate a situation where the navigational style has a clear advantage; participants found the brace and vertical line notations significantly faster than the UNION keyword. The second goal was to determine any differences between various alternative property path constructs. Here the brace notation was significantly faster than the vertical line. The brace notation was also faster than the use of plus, although not significantly so on a pairwise comparison. The advantage of brace may well be that it is an obvious and easily understood notation which enables succinct expression of a query in a rapidly comprehensible form. The final goal was to investigate inversion and negation, and their interaction. Here, the indication is that thinking about either of them is hard. Thinking about both of them at the same time is very hard. This is particularly illustrated by the property path question combining inverse and negation in Section 7, where all the proposed solutions were answered less well than chance. Whilst the difficulties of inverse and negation are likely to be at root cognitive, they may be exacerbated by the non-intuitive symbolism used. Some property graph languages, e.g. Cypher [16], use forward and backward arrows to indicate the direction of a predicate, and this might be helpful for SPARQL. Adapting this notation to our context, *?x friendOf ?y* could be written *?x friendOf -> ?y*, whilst *?x ^friendOf ?y* could be written *?x <- friendOf ?y*. Johnson-Laird [17] describes the American philosopher C.S. Peirce's categorization of signs into: iconic, where representation depends on structural similarity; indexical, where representation depends on a physical connection; and symbolic, where representation depends on convention. The use of ^ is clearly symbolic, whilst the use of arrow is iconic. On the other hand, negation is generally represented symbolically. However, the exclamation mark may not for some people be associated with negation, and a more obvious usage, e.g. *not*, might be helpful. Finally, the analysis of Section 10 is evidence of the benefits offered by graphical representations. Previous work, e.g. [18], suggests that people have a preference either for textual or graphical reasoning. Our study indicates that, at least when thinking about graph databases, the graphical representation is a useful complement of the textual representation.

This leads us to four specific recommendations:

1. Query authors should use predicate paths with vertical line, or better the brace notation (if it is available) in preference to UNION. Where possible, they should minimise their use of negation and inverse, and in particular avoid using these two in combination.
2. Future developments of SPARQL should use more intuitive symbolism. In particular, an arrow notation could be used to represent directionality, in place of, or as an alternative to the use of ^; and *not* used as an alternative to !.
3. The next SPARQL standard should include the braces notation in property paths. In general, query languages should enable succinct and rapidly comprehensible queries, avoiding the need for verbosity.
4. SPARQL query engines should integrate with RDF visualization to support human reasoning about RDF knowledgebases, and in particular to support explanation of query engine results; this could be particularly useful with navigational queries.

**Acknowledgements** The authors would like to thank all study participants; and also Enrico Daga for initial suggestions and assisting with participant contacts.

## References


[1] E. Prud'hommeaux and A. Seaborne, 'SPARQL Query Language for RDF', 2008. https://www.w3.org/TR/rdf-sparql-query/ (accessed Jun. 28, 2019).

[2] S. Harris and A. Seaborne, 'SPARQL 1.1 Query Language', 2013. https://www.w3.org/TR/sparql11-query/ (accessed Jun. 28, 2019).

[3] M. A. Gallego, J. D. Fernández, M. A. Martínez-Prieto, and P. de la Fuente, 'An empirical study of real-world SPARQL queries', 2011, Accessed: Nov. 02, 2015. [Online].

[4] L. Rietveld and R. Hoekstra, 'Man vs. machine: Differences in SPARQL queries', 2014.

[5] A. Bielefeldt, J. Gonsior, and M. Krötzsch, 'Practical Linked Data Access via SPARQL: The Case of Wikidata', 2018.

[6] A. Bonifati, W. Martens, and T. Timm, 'An analytical study of large SPARQL query logs', *VLDB J.*, pp. 1–25, 2017.

[7] P. Warren and P. Mulholland, 'Using SPARQL–the practitioners' viewpoint', in *European Knowledge Acquisition Workshop*, 2018, pp. 485–500.

[8] P. Reisner, 'Human factors studies of database query languages: A survey and assessment', *ACM Comput. Surv. CSUR*, vol. 13, no. 1, pp. 13–31, 1981.

[9] M. K. Sarker, A. Krisnadhi, D. Carral, and P. Hitzler, 'Rule-Based OWL Modeling with ROWLTab Protégé Plugin', in *European Semantic Web Conference*, 2017, pp. 419–433.

[10] P. Warren, P. Mulholland, T. Collins, and E. Motta, 'Improving comprehension of knowledge representation languages: A case study with Description Logics', *Int. J. Hum.-Comput. Stud.*, vol. 122, pp. 145–167, 2019.

[11] A. Seaborne, 'SPARQL 1.1 Property Paths', 2010. https://www.w3.org/TR/sparql11-property-paths/ (accessed Jun. 28, 2019).

[12] M. Arenas, S. Conca, and J. Pérez, 'Counting beyond a Yottabyte, or how SPARQL 1.1 property paths will prevent adoption of the standard', in *Proceedings of the 21st international conference on World Wide Web*, 2012, pp. 629–638.

[13] R Core Team, *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013*. ISBN 3-900051-07-0, 2014.

[14] S. Khemlani, I. Orenes, and P. N. Johnson-Laird, 'Negation: A theory of its meaning, representation, and use', *J. Cogn. Psychol.*, vol. 24, no. 5, pp. 541–559, 2012.

[15] I. Orenes, L. Moxey, C. Scheepers, and C. Santamaría, 'Negation in context: Evidence from the visual world paradigm', *Q. J. Exp. Psychol.*, vol. 69, no. 6, pp. 1082–1092, 2016.

[16] I. Robinson, J. Webber, and E. Eifrem, *Graph databases*. O'Reilly Media, Inc., 2013.

[17] P. N. Johnson-Laird, 'The history of mental models', in *Psychology of reasoning: Theoretical and historical perspectives*, K. Manktelow and M. Chung, Eds. Psychology Press, 2004, p. 179.

[18] M. Ford, 'Two modes of mental representation and problem solution in syllogistic reasoning', *Cognition*, vol. 54, no. 1, pp. 1–71, 1995.